REMARKS

Claims 1-40, 42, and 43 are currently pending in the subject application and are presently under consideration. Claims 2-4, 10, 11, 14, 19, 22, 25, and 33 have been amended as shown on pages 2-10 of the Reply.

Applicant's representative thanks Examiner Hicks for the courtesies extended during the telephonic interview conducted on August 18, 2008. During the discussion, the participants discussed respective interpretations of cited reference Joachims, particularly in connection with the feature of storing unselected but higher-ranked search results to a "non-relevant" training data set. The Examiner helpfully provided his interpretation of the subject claims in view of the cited reference. The remarks below take the Examiner's comments into consideration.

Favorable reconsideration of the subject patent application is respectfully requested in view of the comments and amendments herein.

I.   **Rejection of Claims 1-6, 8-16, 18-22, 29-40, and 42-43 Under 35 U.S.C. §102(a)**

Claims 1-6, 8-16, 18-22, 29-40, and 42-43 stand rejected under 35 U.S.C. §102(a) as being anticipated by Joachims. It is respectfully submitted that this rejection should be withdrawn for at least the following reasons. Joachims does not teach or suggest all features set forth in the subject claims.

> For a prior art reference to anticipate, 35 U.S.C. §102 requires that "each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference." *In re Robertson*, 169 F.3d 743, 745, 49 USPQ2d 1949, 1950 (Fed. Cir. 1999) (*quoting Verdegaal Bros., Inc. v. Union Oil Co.*, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987)).

The subject claims relate to refinement of search query results from a general-purpose search engine based in part on the entry point through which the search engine was accessed. When the search engine is accessed *via* an entry point and a search query is executed, the search query results obtained by the search engine can be passed to a tuning component associated with the entry point. The tuning component can filter and rank the search query results according to a statistical analysis that utilizes two distinct sets of training data associated with the entry point and the context of the search: a first set of data expressly defined as relevant to the search

context and a second data set expressly defined as non-relevant to the search context. The first and second data set can be automatically trained based on observance of user selections. Specifically, when a user selects one of the filtered search query results presented by the tuning component, that result can be automatically recorded by the training component as relevant to the search context and added to the first data set, while results that had been ranked higher than the selected result can be automatically added to the category of non-relevant data. In particular, independent claim 1 recites, *user selection of a query result from a ranked list of the query results causes the selected result to be added to the first set of data and causes the results not selected by the user but ranked higher than the selected result to be automatically added to the second set of data, the first and second sets of data persisted to a computer-readable storage medium.*

Contrary to the Examiner's assertions, Joachims does not disclose such a technique for maintaining sets of training data. Joachims relates to optimization of search result rankings through analysis of recorded clickthrough data. During a search session, clickthrough data is recorded in a logfile for analysis. This clickthrough data is recorded in the form of triplets consisting of a user query (q), the ranking of results (r) presented to the user in response to the query, and the links subsequently selected by the user (c). Once recorded, the clickthrough data is used to train a retrieval function intended to improve search result rankings based on ranking preferences inferred from the logfile. However, the cited reference does not disclose that *selection of a query result can cause non-selected but higher ranked results to be added to a "non-relevant" training data set*, since the training data employed in Joachims' approach does not include both a first data set categorized as "relevant" *and a second data set categorized as "non-relevant."* Indeed, Joachim expressly declares that such a "binary classification" runs counter to the approach explored in that reference. The first paragraph of Section 4, for example, states, "Most work on machine learning in information retrieval does not consider the formulation of above, but simplifies the task to a binary classification problem with the two classes 'relevant' and 'non-relevant'. Such a simplification has several drawbacks…Therefore, the following algorithm directly addresses taking an empirical risk minimization approach." Moreover, the same paragraph concedes that such a data set of non-relevant results *cannot be derived from the clickthrough training data* utilized by Joachims, stating that "such absolute relevance judgments [that is, classifying a result to be non-relevant] *cannot be extracted from*

12

clickthrough data." Clearly, Joachims expressly foregoes the use of "relevant" and "non-relevant" training data sets, and instead attempts to infer relative degrees of relevance from the aforementioned clickthrough triplet data. Since Joachims admittedly does not utilize a training data set of "non-relevant" results, it cannot be said that Joachims in any way teaches or suggests that selection of a search result can cause non-selected but higher ranked results to be added to such a list.

Contending that Joachims discloses these technique for populating a training set of non-relevant data, the Examiner notes that selection of a link from a search result set generated using Joachims' approach causes the clickthrough data for that selection to be recorded, and that this recorded data includes those results not selected by the user. However, the cited reference teaches that this clickthrough data is recorded in the triplet format discussed above (Section 5.1, second paragraph: "The clicks of the user are recorded using the proxy system described in Section 2.1," wherein Section 2.1 discloses recording the *clickthrough data triplets*). As already noted, this data includes the query (q), the ranking (r) presented to the user, and the links selected by the user (c). Presumably, the results not selected by the user are recorded in the ranking (r) information, together with the results that *were* selected by the user. Such a record does not constitute a separate set of data categorized as non-relevant, since this data is only recorded as part of the *total* ranking of results presented to the user. By contrast, the subject claims teach that selection of a search result can cause *both distinct lists* (relevant and non-relevant) to be populated with new training data *via* the techniques disclosed in independent claim 1. Joachims, which expressly foregoes such a binary classification of training data, in no way teaches or suggests these techniques.

Similarly, independent claim 13 recites, *a user clicking a link associated with a search result from the sorted results causes the result to be added to the first set of data and causes the results whose links were not clicked by the user but that are ranked higher than the clicked result to be automatically added to the second set of data.* As discussed *supra*, Joachims fails to teach or suggest this technique for populating a set of relevant and a set of non-relevant training data.

Likewise, independent claim 22 recites, *ranking the filtered general-purpose search engine results; automatically storing a first query result selected by a user in a first data set categorized as relevant; automatically storing at least one non-selected query result that is ranked higher than the first query result in a second data set categorized as non-relevant upon*

selection of the first query result. These aspects are not disclosed in Joachims, as already discussed.

Independent claim 29 discloses similar aspects, reciting, *selecting a link associated with a query result from the list; automatically adding the selected query result to the first set of data; and automatically adding non-selected results from the list that are ranked higher than the selected query result to the second set of data upon selection of the selected query result.* Joachims is silent regarding these aspects, as discussed above.

Moreover, independent claim 34 recites, *recording a first query result from a ranked list of query results returned from the executed query as relevant when a user views the document associated with the first query result; recording at least one second query result whose associated document was not viewed by the user but that is ranked higher than the first query result as non-relevant when the first result is selected for viewing by the user,* and as already discussed, the cited reference is silent regarding these aspects.

Independent claim 42 recites, *selecting a link associated with a first search result from the ranked results causes the first result to be added to the first set of data and causes results that are ranked higher than the first result and have not been selected by the user to be automatically added to the second set of data.* The cited reference does not disclose these techniques for collecting training data, as noted above.

Disclosing similar features, independent claim 43 recites, *a user viewing a document associated with a first search result from the ranked results causes the first result to be added to the first set of training data and causes the results that are unviewed but ranked higher than the first result to be automatically added to the second set of training data.* As discussed *supra,* Joachims fails to teach or suggest these features.

In view of at least the foregoing, it is respectfully submitted that Joachims does not teach or suggest each and every feature of independent claims 1, 13, 22, 29, 34, 42, and 43 (and all claims depending there from), and as such fails to anticipate the subject invention. It is therefore requested that this rejection be withdrawn.

II.   Rejection of Claims 7, 17, and 23-28 Under 35 U.S.C. §103(a)

Claims 7, 17, and 23-28 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Joachims in view of Pazzani, *et al.* It is respectfully submitted that this rejection should be withdrawn for at least the following reasons. Joachims and Pazzani, *et al.*, individually or in combination, do not teach or suggest each and every feature set forth in the subject claims.

> A factfinder should be aware, of course, of the distortion caused by hindsight bias and must be cautious of arguments reliant upon *ex post* reasoning. See *KSR v. Teleflex*, 550 U.S. ___, 127 S. Ct. 1727 (2007) citing Graham v. John Deere Co. of Kansas City, 383 U. S. 1, 36 (warning against a "temptation to read into the prior art the teachings of the invention in issue" and instructing courts to "'guard against slipping into the use of hindsight'" (*quoting Monroe Auto Equipment Co. v. Heckethorn Mfg. & Supply Co.*, 332 F. 2d 406, 412 (CA6 1964))).

Claim 7 depends from independent claim 1, claim 17 depends from independent claim 13, and claims 23-28 depend from independent claim 22. As discussed in the previous section of the Reply with respect to those independent claims, Joachims does not teach or suggest that selection of a search result can cause the selected result to be added to a training data set of relevant results, while causing non-selected but higher ranked search results to be added to a data set of non-relevant results. Pazzini, *et al.* is also silent regarding this manner of collecting training data. Pazzini, *et al.* presents an overview of an intelligent agent called Syskill & Webert, which is used to develop and refine user profiles that infer websites of interest to the associated user. These user profiles can be revised and updated in response to feedback from the associated user regarding which websites are of interest and which are not, and these updated profiles can be used to predict which websites will be of most interest to the user. Although Pazzini, *et al.* teaches that algorithms used to update the user profiles employ a set of positive examples (*e.g.* websites of interest to the user) and negative examples (*e.g.* websites the user is not interested in), the cited reference indicates that these examples must be *explicitly selected* by the user in both cases. Section 2.1, paragraph 3, for example, explains that pages are rated by a user as being "hot" (interesting), or "cold" (not interesting), and these ratings are used to train the algorithm that refines the user profiles. Hence, websites that a user finds uninteresting for a given search session must be explicitly selected by the user for inclusion in the negative examples. The subject claims, by contrast, disclose that query results ranked higher than a

selected result can automatically be included in the set of non-relevant training data without the need to visit or rank the non-relevant sites, thereby removing the burden of explicitly indicating uninteresting websites from the user. Pazzini, *et al.* does not disclose such an automated technique for collecting training data.

In view of at least the foregoing, it is respectfully submitted that Joachims, alone or in combination with Pazzini, *et al.*, does not teach or suggest all aspects set forth in independent claims 1, 13, and 22 (and all claims depending there from), and as such fails to make obvious the subject invention. It is therefore requested that this rejection be withdrawn.

CONCLUSION

The present application is believed to be in condition for allowance in view of the above comments and amendments. A prompt action to such end is earnestly solicited.

In the event any fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063 [MSFTP444US].

Should the Examiner believe a telephone interview would be helpful to expedite favorable prosecution, the Examiner is invited to contact applicant's undersigned representative at the telephone number below.

Respectfully submitted,

AMIN, TUROCY & CALVIN, LLP

  /Himanshu S. Amin/
Himanshu S. Amin
Reg. No. 40,894

AMIN, TUROCY & CALVIN, LLP
24TH Floor, National City Center
1900 E. 9TH Street
Cleveland, Ohio 44114
Telephone (216) 696-8730
Facsimile (216) 696-8731